(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENT- UND MARKENAMT

(12) **Offenlegungsschrift**

(10) **DE 198 40 548 A 1**

(51) Int. Cl.[7]: **G 10 L 11/00**

(21) Aktenzeichen: 198 40 548.0
(22) Anmeldetag: 27. 8. 1998
(43) Offenlegungstag: 2. 3. 2000

DE 198 40 548 A 1

(71) Anmelder:
Deutsche Telekom AG, 53113 Bonn, DE

(72) Erfinder:
Berger, Jens, 10405 Berlin, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht zu ziehende Druckschriften:

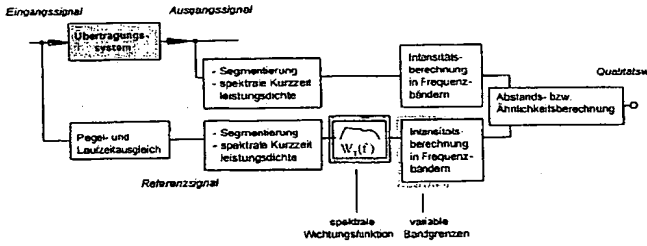
DE 37 08 002 A1
EP 08 09 236 A1
EP 07 27 767 A2

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gem. § 44 PatG ist gestellt

(54) Verfahren zur instrumentellen ("objektiven") Sprachqualitätsbestimmung

(57) Bekannte Verfahren zur instrumentellen Sprachqualitätsbestimmung auf der Basis eines Vergleichs von Signalintensitäten des zu bewertenden Sprachsignals mit einem Referenzsprachsignal bewerten spektrale Verformungen des zu bewertenden Sprachsignals nicht optimal, so daß die Qualitätsbewertung unsicher ist. Des weiteren werden durch die Integration der Signalintensität in Frequenzbändern mit konstanten Bandgrenzen bestimmte Verfälschungen des zu bewertenden Sprachsignals, wie sie z. B. durch Codiersysteme niederer Bitraten hervorgerufen werden, fehlerhaft bewertet.
Um die Aussagesicherheit der berechneten Qualitätskennwerte zu erhöhen, werden zum einen Verformungen der mittleren spektralen Einhüllenden vor einem Vergleich der spektralen Eigenschaften mit einer Wichtungsfunktion $W_T(f)$ weitgehend korrigiert. Zum anderen werden die festen Bandgrenzen zur Integration der spektralen Leistungsdichte aufgehoben und statt dessen in einem vorgegebenen Optimierungsbereich Bandgrenzen gesucht, bei denen die sich ergebenden spektralen Intensitätsabbildungen von zu bewertenden Sprachsignal und Referenzsprachsignal eine maximale Ähnlichkeit aufweisen.
Die beschriebenen Lösungen können bekannte Verfahren erweitern und zu deren Struktur hinzugefügt werden.



DE 198 40 548 A 1

Beschreibung

Vorbemerkung

Die Erfindung bezieht sich auf ein Verfahren zur instrumentellen ("objektiven") Sprachqualitätsbestimmung, bei dem durch Vergleich von Eigenschaften eines zu bewertenden Sprachsignals mit Eigenschaften eines Referenzsprachsignals (ungestörtes Signal) Kennwerte zur Bestimmung der Sprachqualität (Sprachgüte) abgeleitet werden.

Sprachqualitätsbestimmungen von Sprachsignalen werden in der Regel mittels auditiver ("subjektiver") Untersuchungen mit Versuchspersonen vorgenommen.

Das Ziel von instrumentellen ("objektiven") Verfahren zur Sprachqualitätsbestimmung ist es, aus Eigenschaften des zu bewertenden Sprachsignals mittels geeigneter Rechenverfahren Kennwerte zu ermitteln, die die Sprachqualität des zu bewertenden Sprachsignals beschreiben, ohne auf Urteile von Versuchspersonen zurückgreifen zu müssen.

Die berechneten Kennwerte und das zugrunde gelegte Verfahren zur instrumentellen Sprachqualitätsbestimmung gelten als anerkannt, wenn eine hohe Korrelation zu Ergebnissen auditiver Vergleichsuntersuchungen erreicht wird. Die mittels auditiver Untersuchungen gewonnenen Sprachqualitätswerte stellen somit die Zielwerte dar, die durch instrumentelle Verfahren erreicht werden sollen.

Stand der Technik

Bekannte Verfahren zur instrumentellen Sprachqualitätsbestimmung beruhen auf einem Vergleich eines Referenzsprachsignals mit dem zu bewertenden Sprachsignal. Dabei werden das Referenzsprachsignal und das zu bewertendes Sprachsignal in kurze Zeitabschnitte segmentiert. In diesen Segmenten werden die spektralen Eigenschaften der beiden Signale verglichen.

Für die Berechnung der spektralen Kurzzeiteigenschaften kommen verschiedene Ansätze und Modelle zur Anwendung. In der Regel erfolgt die Berechnung der Signalintensität in Frequenzbändern, deren Breite mit zunehmender Mittenfrequenz größer wird. Beispiele für solche Frequenzbänder sind die bekannten Terzbänder oder Frequenzgruppen nach Zwicker (veröffentlicht in Zwicker, E.: "Psychoakustik", Berlin: Springer-Verlag, 1982).

Die derart berechnete spektrale Intensitätsabbildung für jeden betrachteten Zeitabschnitt läßt sich als Reihe von Zahlenwerten auffassen, in der die Anzahl der Einzelwerte der Anzahl der verwendeten Frequenzbänder entspricht, die Zahlenwerte selbst die berechneten Intensitätswerte darstellen und ein fortlaufender Index der Frequenzbänder die Reihenfolge der Zahlenwerte beschreibt.

Bei den derzeit bekannten Verfahren zur instrumentellen Sprachqualitätsbestimmung werden die Grenzen der benutzten Frequenzbänder auf der Frequenzachse konstant gehalten.

In jedem betrachteten Zeitsegment werden die berechneten Intensitäten von zu bewertenden Sprachsignal und Referenzsprachsignal in jedem Band miteinander verglichen. Die Differenz beider Werte, bzw. die Ähnlichkeit der beiden entstehenden spektralen Intensitätsabbildungen, stellt die Grundlage für die Berechnung eines Qualitätswertes dar (**Fig.** 1).

Solche Verfahren wurden insbesondere für die qualitative Bewertung der Sprache in der Telefonieanwendung entwikkelt. Beispiele hierfür sind die Veröffentlichungen:

"A perceptual speech-quality measure based on a psychacoustic sound representation" (Beerends, J. G.: Stemerdink. J. A., J. Audio Eng. Soc. 42 (1994) 3, S. 115–123).

"Auditory distortion measure for speech coding" (Wang, S. Sekey, A.; Gersho, A.: IEEE Proc. Int. Confacoust., speech and signalprocessing (1991), S. 493–496).

Der derzeit gültige ITU-T Standard P.861 beschreibt ebenfalls ein derartiges Verfahren: "Objective quality measurement of telephone-band speech codecs" (ITU-T Rec. P.861, Genf 1996).

Nachteile bekannter instrumenteller Sprachqualitätsmeßverfahren

Der Einsatz von bekannten Verfahren zur instrumentellen Sprachqualitätsbestimmung scheitert an der Zuverlässigkeit der berechneten Qualitätswerte für bestimmte zu bewertende Signaleigenschaften. Insbesondere bei Beeinträchtigungen im zu bewertenden Sprachsignal, wie sie z. B. durch Sprachcodierverfahren mit niedrigen Bitraten oder Kombinationen von unterschiedlichen Störungen hervorgerufen werden, liefern derzeit bekannte Verfahren nur unsichere Qualitätswerte.

Nachteilig bei den heute bekannten Verfahren ist in solchen Fällen, daß bei einem Vergleich zwischen dem zu bewertenden Sprachsignal mit einem Referenzsprachsignal Unterschiede zwischen beiden Signalabschnitten in der gewählten Darstellungsebene in den zu berechnenden Qualitätskennwert einfließen, die nicht oder kaum zu einer – auch im auditiven Test wahrnehmbaren – qualitativen Beeinträchtigung führen.

Im Rahmen der hier betrachteten Sprachübertragung in Telefonanwendungen tragen Frequenzbandbegrenzungen und spektrale Verformungen des zu bewertenden Sprachsignals (z. B. hervorgerufen durch Filtereigenschaften des Telefongerätes oder des Übertragungskanals) nur begrenzt zu einer empfundenen qualitativen Beeinträchtigung bei.

Um diese Mängel teilweise zu vermeiden, wird in einem anderen Ansatz versucht, die linearen Verzerrungen (Frequenzgang) durch ein Korrekturfilter bzw. eine Leistungsübertragungsfunktion zu kompensieren (veröffentlicht in: "A new approach to objective quality-measures based on attribute-matching", Halka, U.; Heute, U., Speech communication, 11 (1992) 1, S. 15–30). Die Anwendung dieses Verfahrens ist jedoch bei nichtlinearer und zeitinvarianter Übertragung nachteilig, da die so berechnete Kompensationsfunktion nicht mehr ausschließlich die spektralen Verformungen des zu bewertenden Signals beschreibt.

Verschiebungen spektraler Kurzzeit-Maxima ("Formantverschiebungen") im zu testenden Signal gegenüber dem Referenzsprachsignal, z. B. verursacht durch Codiersysteme mit niedriger Bitrate, führen bei bekannten Verfahren zu großen Unterschieden in den spektralen Intensitätsabbildungen und gehen damit stark in den berechneten Qualitätswert ein. Untersuchungen haben ergeben, daß in einer auditiven Sprachqualitätsuntersuchung diese Verschiebungen spektraler Kurzzeit-Maxima jedoch nur begrenzten Einfluß auf das Qualitätsurteil haben.

Aufgabe

Die Erfindung stellt sich die Aufgabe, den Einfluß von spektralen Begrenzungen und Verformungen des zu bewertenden Sprachsignals sowie von Verschiebungen spektraler Kurzzeit-Maxima vor dem Vergleich der spektralen Eigenschaften eines zu testenden Signals mit einem Referenzsprachsignal und der Berechnung eines Qualitätswertes in instrumentellen Verfahren zu reduzieren.

Lösung

Im Gegensatz zu bekannten Ansätzen wird in der hier beschriebenen Erfindung eine spektrale Wichtungsfunktion generiert, die auf mittleren spektralen Einhüllenden, z. B. der mittleren spektralen Leistungsdichte, von zu bewertendem Sprachsignal und Referenzsprachsignal beruht. Dies ermöglicht den Einsatz des Verfahrens ebenfalls bei nichtlinearer und zeitvarianter Übertragung.

Die spektrale Wichtungsfunktion wird aus den Quotienten der Stützwerte der mittleren spektralen Leistungsdichte des zu bewertenden Signals $Phi_Y(f)$ und der des Eingangssignals des Übertragungssystems $Phi_X(f)$ derart berechnet, daß die Wichtungsfunktion über

$$W_T(f) = a(f) \cdot (Phi_Y(f)/Phi_X(f))$$

zu beschreiben ist. Die Bewertungsfunktion a(f) kann die Wichtungsfunktion $W_T(f)$ an über den Wirkungsbereich unterschiedlich gewichten, sie ist im einfachsten Falle konstant 1.

Die derart berechnete spektrale Wichtungsfunktion $W_T(f)$ nähert die mittleren spektralen Einhüllenden von zu bewertenden Sprachsignal und Referenzsprachsignal einander an, so daß Unterschiede der beiden spektralen Einhüllenden nur noch vermindert in den berechneten Qualitätswert einfließen.

Die spektrale Wichtungsfunktion $W_T(f)$ kann zum einen auf das Referenzsprachsignal angewendet werden. Dabei wird das Referenzsprachsignal in seiner mittleren spektralen Leistungsdichte dem zu bewertenden Signal angenähert (**Fig.** 2a).

Zum anderen kann die spektrale Wichtungsfunktion invertiert auf das zu bewertende Signal angewendet werden. Dieses wird dadurch entzerrt und, hinsichtlich seiner mittleren spektralen Leistungsdichte, an das Referenzsprachsignal angenähert (**Fig.** 2b).

Ein weiterer Teil der Erfindung bezieht sich auf die Korrektur von Verschiebungen spektraler Kurzzeit-Maxima, die durch die Übertragungssysteme verursacht werden.

Die Intensität wird für jeden Zeitabschnitt in Frequenzbändern integriert. Resultat ist eine Reihe von Intensitätswerten für jede spektrale Darstellung eines Signalabschnitts, wobei jeder Einzelwert die Intensität in einem Frequenzband repräsentiert. Die Verschiebungen spektraler Kurzzeit-Maxima können hierbei zu abweichenden berechneten Intensitäten in den Frequenzbändern von Referenzsprachsignal und zu bewertenden Sprachsignal führen.

Diese Abweichungen in den spektralen Intensitätsabbildungen – verursacht Verschiebungen spektraler Kurzzeit-Maxima – können durch eine variable Anordnung der Frequenzbänder auf der Frequenzachse reduziert werden. Im Gegensatz zu den konstanten Bandgrenzen bei bekannten Verfahren werden die Bandgrenzen auf der Frequenzachse verschoben. Die Zahl der Frequenzbänder und deren Index bleibt aber konstant. In einer Optimierungsschleife werden dann diejenigen Bandgrenzen akzeptiert, bei denen die beiden entstehenden spektralen Abbildungen von zu bewertenden Sprachsignal und Referenzsprachsignal maximale Ähnlichkeit aufweisen bzw. deren Abstand minimal ist. Diese Optimierung wird für alle Bänder in allen betrachteten Zeitsegmenten durchgeführt.

Der Einsatz variabler Bandgrenzen zur Berechnung der spektralen Intensitätsabbildung ist nicht nur auf das Signal, in dem auch die beschriebene spektrale Wichtungsfunktion $W_T(f)$ zum Einsatz kommt, beschränkt, sondern kann auch auf das jeweils andere Signal und sogar auf beide Signale angewendet werden. (vgl. **Fig.** 2a und 2b).

Ausführungsbeispiel

Ein spezielles Ausführungsbeispiel zeigt eine Realisierung gemäß **Fig.** 3, die als TOSQA (Telecommunication Objective Speech Quality Assessment) bezeichnet wird. Hierbei erfolgt eine erweiterte Vorverarbeitung des Referenzsprachsignals.

In Spezifikation der allgemeinen Realisierungen nach **Fig.** 2a und 2b werden hier Sprachpausen mittels eines Sprachpausenerkenners erkannt und gehen nicht in das Qualitätsmaß ein. Ebenfalls erfolgt eine Filterung von Referenzsprachsignal und zu bewertendem Sprachsignal mit einem Bandpaß 300 . . . 3400 Hz sowie eine Filterung auf den Frequenzgang eines Telefonhandapparates. Die Integration der spektralen Leistungsdichte erfolgt in Frequenzgruppen, die die Basis für die Berechnung der spezifischen Lautheit darstellen.

Die Integration in Frequenzgruppen erfolgt jedoch nicht in festen Frequenzgruppengrenzen, sondern mit den in dieser Erfindung beschriebenen variablen Frequenzgruppengrenzen. Die berechneten Signalleistungen in den so modifizierten Frequenzgruppen bilden die Basis für die Intensitätsberechnung. Hier wurde auf ein Modell zur Berechnung der spezifischen Lautheit nach Zwicker, einer gehörrichtigen Intensitätsabbildung, zurückgegriffen (veröffentlicht in Zwikker, E.: "Psychoakustik", Berlin: Springer-Verlag, 1982).

Die berechneten Lautheitsmuster werden in Ergänzung des allgemeinen Ansatzes noch durch eine Fehlerbewertungsfunktion ergänzt. Der berechnete Qualitätswert wird über einen Mittelwert der Korrelationskoeffizienten der spezifischen Lautheiten für jedes betrachtete kurze Zeitsegment über die Zahl der ausgewerteten Sprachsegmente gebildet.

Patentansprüche

1. Verfahren zur instrumentellen Sprachqualitätsbestimmung, bei dem durch Vergleich von spektralen Kurzzeiteigenschaften eines zu bewertenden Sprachsignals mit einem Referenzsprachsignal Kennwerte zur Bestimmung der Sprachqualität berechnet werden, **dadurch gekennzeichnet**, daß vor dem Vergleich der Eigenschaften der Sprachsignale, Unterschiede in mittleren spektralen Einhüllenden verringert werden, indem aus diesen zuerst eine spektrale Wichtungsfunktion berechnet wird, mit der die spektralen Kurzzeiteigenschaften der Sprachsignale in allen betrachteten Zeitsegmenten gewichtet werden, so daß die Unterschiede in den mittleren spektralen Einhüllenden dadurch nur begrenzt in den zu berechnenden Qualitätskennwert einfließen, und daß für die Berechnung der Signalintensität die Grenzen der benutzten Frequenzbänder variabel gestaltet werden, so daß für jeden betrachteten Signalabschnitt in jeweils allen ausgewerteten Frequenzbändern die berechneten Intensitäten von Referenzsprachsignal und zu bewertendem Signal zueinander möglichst geringe Unterschiede aufweisen.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß zuerst die mittleren spektralen Einhüllenden von zu bewertenden Sprachsignal und Referenzsprachsignal in Form eines mittleren Leistungsdichtespektrums berechnet werden und aus dem Quotienten beider Spektren eine spektrale Wichtungsfunktion $W_T(f)$ berechnet wird, mit der die Kurzzeit-Leistungsdichtespektren des Referenzsprachsignals vor der Berechnung eines Qualitätskennwertes gewichtet werden.
3. Verfahren nach Anspruch 1 und 2, dadurch gekennzeichnet, daß die zu berechnende Wichtungsfunktion $W_T(f)$ nur aus Teilbereichen der berechneten mittleren

spektralen Einhüllenden von zu bewertenden Sprachsignal und Referenzsprachsignal berechnet wird und damit die Unterschiede in mittleren spektralen Einhüllenden zwischen beiden Signalen nur in spektralen Teilbereichen verringert werden.

4. Verfahren nach Anspruch 1 bis 3, dadurch gekennzeichnet, daß vor Berechnung der Qualitätskennwerte eine Integration der Signalintensität für jeden ausgewerteten kurzen Zeitabschnitt in Frequenzgruppen erfolgt, wobei die Grenzen der Frequenzgruppen auf der Frequenzachse variabel sind, aber die Breite der Frequenzgruppen auf der Tonheitskala konstant bleibt, und daß aus den Signalintensitäten in den Frequenzgruppen eine Berechnung der spezifischen Lautheit erfolgt, wobei die Grenzen der Frequenzgruppen benutzt werden, bei denen die berechneten Unterschiede in der spezifischen Lautheit zwischen dem zu bewertenden Signal und dem Referenzsprachsignal im jeweils betrachteten Band und Zeitsegment den geringsten Unterschied aufweisen.

5. Verfahren nach Anspruch 1 bis 4, dadurch gekennzeichnet, daß der Qualitätskennwert aus der Ähnlichkeit der spektralen Darstellungen in jedem betrachteten Zeitabschnitt berechnet wird, wobei die Ähnlichkeit einen über alle betrachteten Zeitabschnitte gemittelten Korrelationskoeffizienten zwischen der spektralen Darstellung des zu bewertenden Sprachsignals und der spektralen Darstellung des Referenzsprachsignals im jeweiligen Zeitsegment darstellt.

6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, daß der Korrelationskoeffizient zwischen der spektralen Darstellung des zu bewertenden Sprachsignals und der spektralen Darstellung des Referenzsprachsignals im jeweiligen Zeitsegment nur von einem Teilbereich der spektralen Darstellung berechnet wird, d. h. für die Berechnung des Qualitätskennwertes nicht alle berechneten Spektralwerte berücksichtigt werden.

Hierzu 4 Seite(n) Zeichnungen

Nummer: DE 198 40 548 A1
Int. Cl.[7]: G 10 L 11/00
Offenlegungstag: 2. März 2000

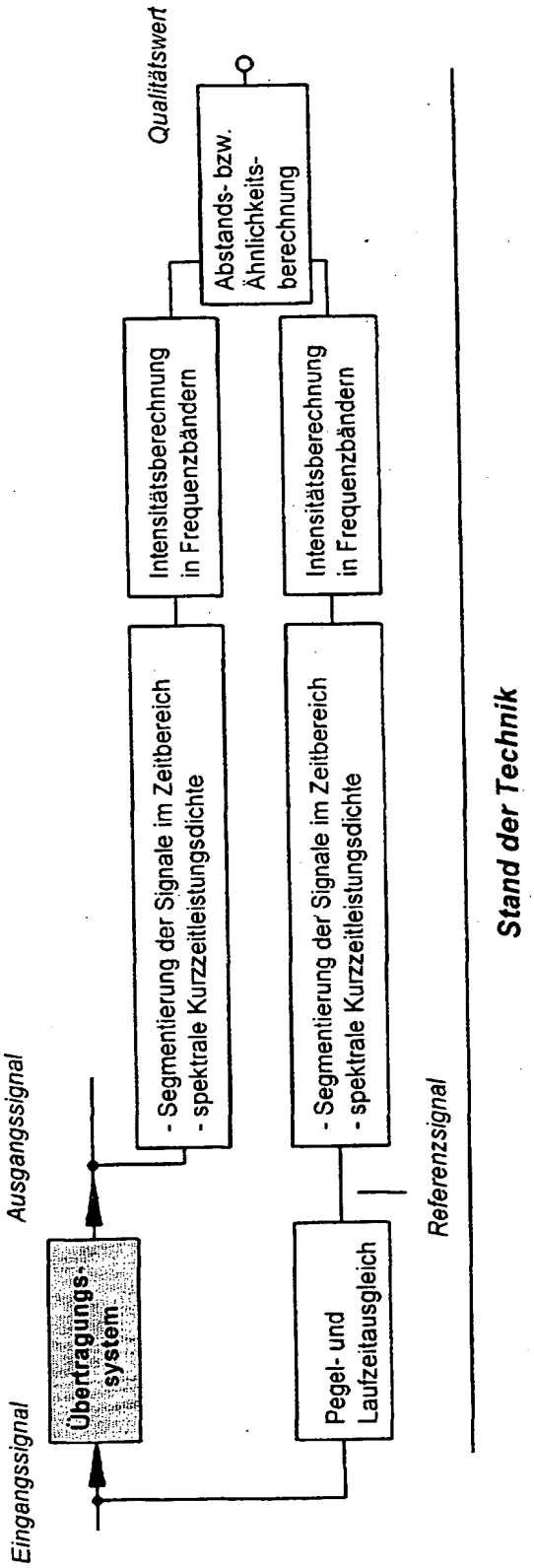


Fig. 1

Nummer: DE 198 40 548 A1
Int. Cl.[7]: G 10 L 11/00
Offenlegungstag: 2. März 2000

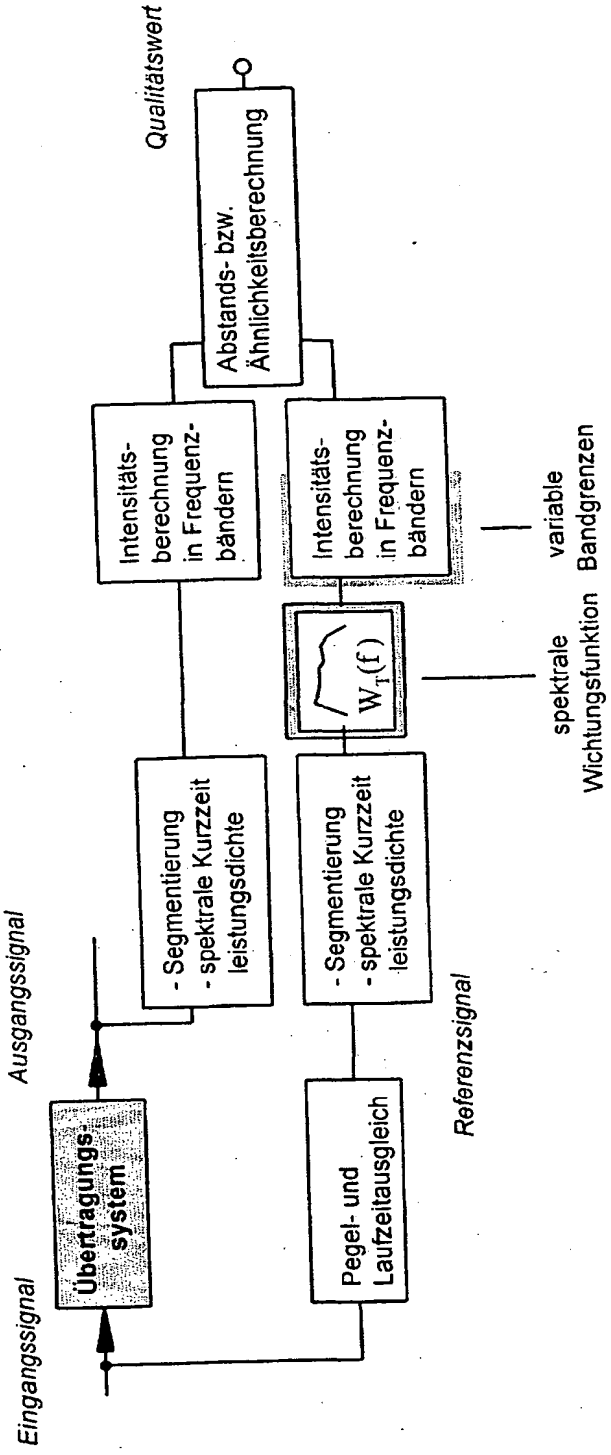


Fig. 2a

Nummer: DE 198 40 548 A1
Int. Cl.[7]: G 10 L 11/00
Offenlegungstag: 2. März 2000

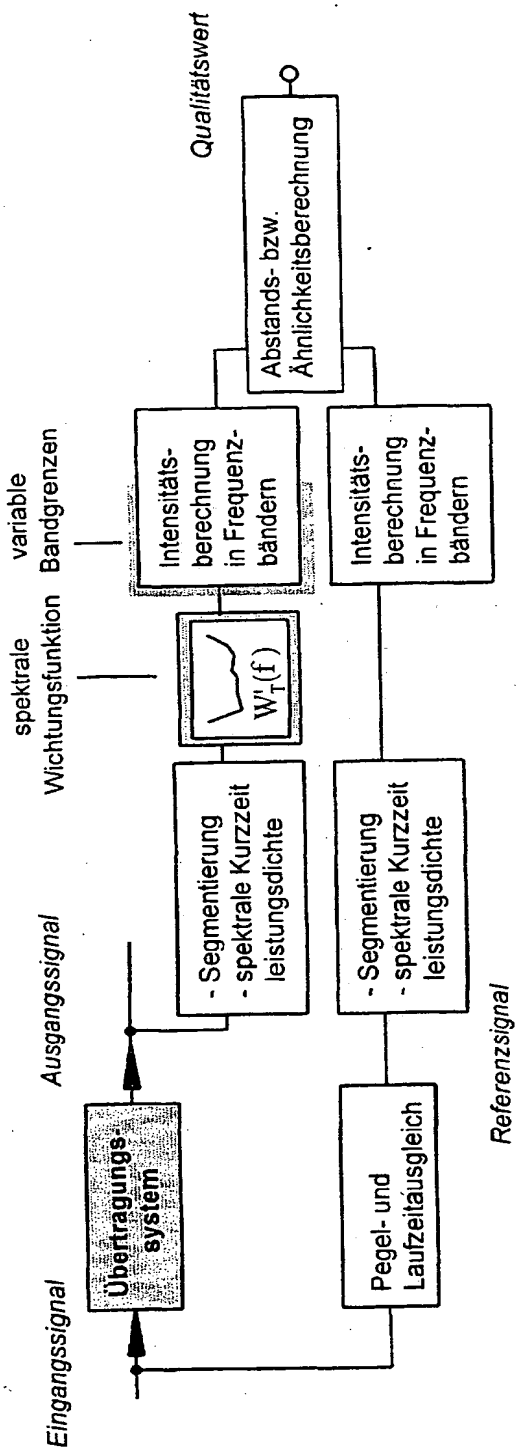


Fig. 2b

Nummer: DE 198 40 548 A1
Int. Cl.[7]: G 10 L 11/00
Offenlegungstag: 2. März 2000

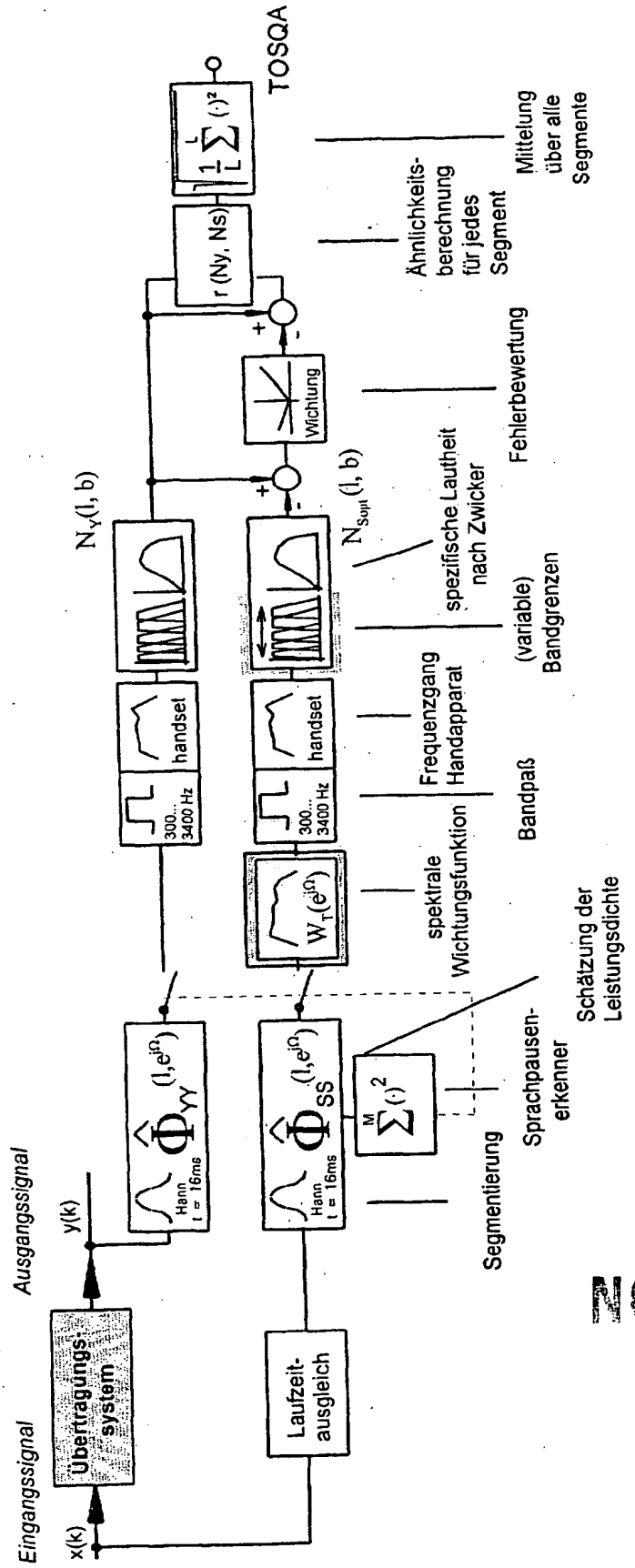


Fig. 3